DOCUMENT RESUME

ED 213 803 UD 022 151

AUTHOR Goldberger, Arthur S.; Cain, Glen G.

TITLE The Causal Analysis of Cognitive Outcomes in the Coleman Report.

INSTITUTION Wisconsin Univ., Madison. Inst. for Research on Poverty.

SPONS AGENCY Department of Health and Human Services, Washington, D.C.; National Inst. of Education (ED), Washington, D.C.; National Science Foundation, Washington, D.C.; Wisconsin Univ., Madison. Graduate School.
REPORT NO IRP-DP-682-81
PUB DATE Dec 81
GRANT BNS-76-22943; NIE-G-81-0009; SES-80-08053
NOTE 58p.; Also supported by funds from the William F. Vilas Trust Estate.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Academic Achievement; Multiple Regression Analysis; *Predictor Variables; *Private Schools; *Research Methodology; *School Policy; Secondary Education; Student Characteristics; Testing
IDENTIFIERS *Public and Private Schools (Coleman et al)

ABSTRACT

In the study "Public and Private Schools," the conclusions by James Coleman and others that private schools produce higher test scores and do so by employing certain school policies are not valid because the methods and interpretations used in the analysis fall below the standards for social-scientific research. Moreover, the conclusions are not warranted by their evidence. In implementing the research, Coleman and his associates used inadequate measures of cognitive outcomes, employed erroneous sampling procedures, and based conclusions on biased methods of analysis. One phase of analysis used a multiple regression approach with 17 background variables that was selectively biased against public schools. A second analysis, which attempted to show that Catholic schools are more egalitarian because they produce similar test scores among students with diverse backgrounds, was biased because it excluded relevant variables and because the range of student cognitive abilities was narrower in the Catholic sector than in the public sector. The third analysis which studied sophomore to senior change in scores did not control for background. Finally, the analysis which attributed test score gains in the private sector to selected school policies was based on choosing certain variables as "policy" without considering student ability and background. (Author/MJL)

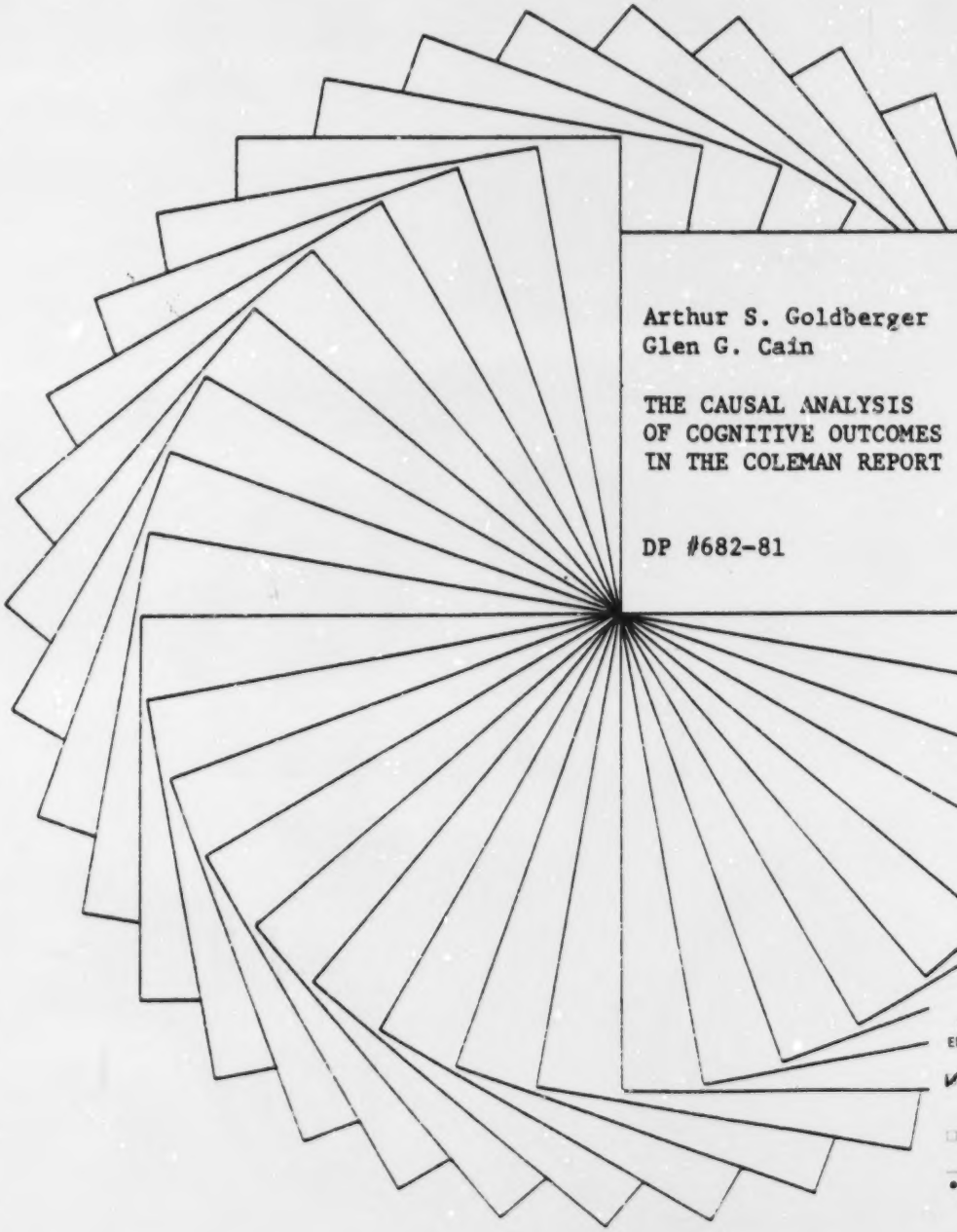
****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

University of Wisconsin-Madison

Institute for Research on Poverty

Discussion Papers

ED213803



Arthur S. Goldberger
Glen G. Cain

THE CAUSAL ANALYSIS
OF COGNITIVE OUTCOMES
IN THE COLEMAN REPORT

DP #682-81

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

UD 022 151

The Causal Analysis of Cognitive Outcomes in the Coleman Report

Arthur S. Goldberger
Department of Economics and
Institute for Research on Poverty,
University of Wisconsin-Madison

Glen G. Cain
Department of Economics and
Institute for Research on Poverty,
University of Wisconsin-Madison

December 1981

Goldberger's work was supported by grants from the National Science Foundation (BNS-76-22943 and SES-80-08053), the University of Wisconsin Graduate School Research Committee, and the William F. Vilas Trust Estate. Cain's work was supported by the National Institute of Education under Grant No. NIE-G-81-0009 to the Wisconsin Center for Education Research and by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the U.S. Department of Health and Human Services pursuant to the Economic Opportunity Act of 1964. The authors are grateful to Lee Cronbach, Robinson Hollister, Richard Murnane, Michael Olneck, and David Rogosa for helpful comments. The opinions, findings, and conclusions expressed are those of the authors and do not necessarily reflect the views of the Institutes or the sponsoring Departments.

ABSTRACT

Cognitive test scores in a large sample of American high school students were analyzed in Public and Private Schools, a report by James Coleman, Thomas Hoffer, and Sally Kilgore. This latest "Coleman Report" came to strong conclusions about the effects of the public, Catholic, and other-private schools: that the private sectors produce higher test scores, and that they do so by employing better school policies.

We provide a detailed critical evaluation of this material. We find that the methods and interpretations employed fall below the minimum standards of acceptability for social-scientific research. We also find that the strong conclusions are not warranted by the evidence.

THE CAUSAL ANALYSIS OF COGNITIVE OUTCOMES IN THE COLEMAN REPORT

1. INTRODUCTION

In their report to the National Center for Education Statistics [1], James Coleman, Thomas Hoffer, and Sally Kilgore (henceforth CHK) address three questions with respect to cognitive outcomes (= test scores) in the American high school system (= 3 sectors: public, Catholic, and other private): (1) Do mean test scores differ across the three sectors? (2) Do the sectors actually produce the differences in outcome? (3) How do the sectors produce those differences -- that is, by what policies do they accomplish their effects? CHK provide these answers: (1) Yes: mean test scores are higher in the private sectors than in the public sector. (2) Yes: test scores are higher in the private sector even after controlling for differences in student characteristics. Furthermore, the private sectors produce larger gains in test scores during the last two years of high school than does the public sector. Furthermore, the Catholic schools come closer to the "common school" ideal, by educating students of varying background more nearly alike than do the other-private and public schools. (3) The private schools produce their higher test scores by placing higher academic demands and imposing stricter discipline on their students than do the public schools.

Our task is to assess the validity of these answers -- that is, to evaluate the evidence and reasoning that generated them. We have been handicapped by the style of the Report, a document of 233 pages + appendices. Elaborate calculations from various regression equations are given, but

the equations themselves are rarely presented. Sample sizes are seldom indicated. The definitions of variables are often cryptic. In such circumstances one might want to rely on the objectivity and scientific judgment of the authors as a substitute for the documentation one expects to find in a scientific report. But there is so much advocacy in the CHK Report that this option was unavailable to us. We have, however, been assisted by access to some of the authors' computer output provided to us on request, and by access to unpublished studies by others who have been reanalyzing the original data set. Finally, we have tapped CHK's article, "Cognitive outcomes in public and private schools" [2], and Coleman's own restatement, "Private schools, public schools, and the public interest" [3].

Our summary assessments are that the methods and interpretations used by CHK fall below the minimum standards for social-scientific research, and that CHK's answers to the questions posed above are not warranted by their evidence.

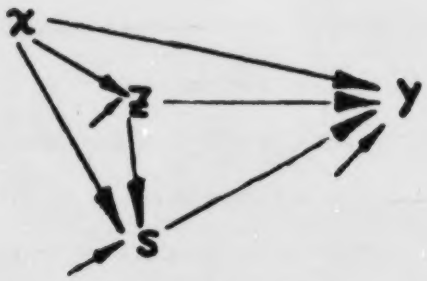
2. MODELS AND DATA

A. Causal Model

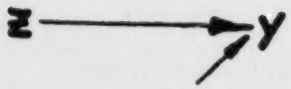
We begin with a formulation of the causal models implicit in CHK's analysis. Figure 1A presents a general scheme for school effects on cognitive outcomes which appears to underlie the text and tables in Chapter 6 of the Report. In our diagrams, x = background, z = sector, s = school policy, y = test score, the straight arrows represent direct causal paths, with the short arrows denoting residual paths, and the curved arrows represent noncausal associations. The first phase of the analysis

Figure 1

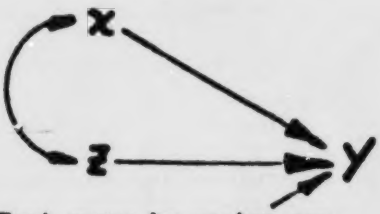
Causal Models for School Sector Effects on Cognitive Achievement



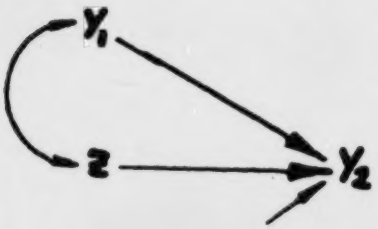
1A: General scheme.



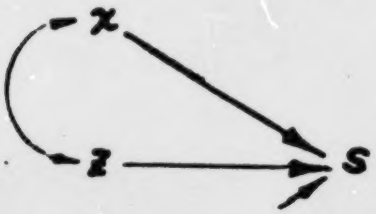
1B: Test score by sector.



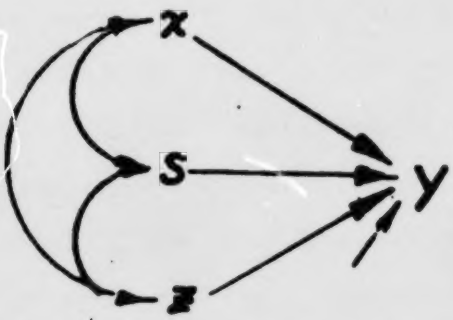
1C: Test score by sector, controlling for background.



1D: Senior test score by sector, controlling for sophomore test score.



1E: School policy by sector, controlling for background.



1F: Test score by sector, controlling for background and school policy.

Symbols: x = background, z = sector, s = school policy, y = test score. Straight arrows denote causal paths, short arrows denote residual paths, curved arrows denote noncausal associations.

[1:154-161] takes up the "raw" association between sector and test score, as indicated in Figure 1B. In the second phase of the analysis [1:165-180], the background-controlled relationship between sector and test score is investigated, as indicated in Figure 1C. (Observe that as compared with Figure 1A, the path from x to z has been collapsed into an association between x and z, a simplification which is justified when residuals leading to z and y in Figure 1A are uncorrelated; that is, when unmeasured determinants of sector choice are unrelated to unmeasured determinants of test score.) The second phase of the analysis [1:180-185] also contains an alternative attempt at controlling for differences between the students, in which, as effectively in Figure 1D, y_2-y_1 (the change in test score from sophomore to senior year) is related to sector. The third phase [1:197-219] starts with an analysis of the relationship of school policy to background and sector (as in Figure 1E) and proceeds to the relation between test score and sector, now controlling for school policy as well as background (as in Figure 1F).

B. Sectors and Test Scores

Mean test scores by sector, along with the standard deviations [1:154, Table 6.1.5; 2: Table 3] are reproduced here as Table 1. (Note: here and subsequently, we refer to both [1] and [2] as sources for the same tabular material. In the event that entries differ between the two sources, we use those from [2]). There are three tests: Reading, Vocabulary, and Mathematics (henceforth, R, V, M) which contain respectively 8, 8, and 18 items. These test scores are the primary dependent variables throughout CHK's study of cognitive outcomes in American high schools. The tests,

Table 1

Mean (and Standard Deviation) of Test Scores,
by Grade and Sector

	Reading		Vocabulary		Mathematics	
	Soph	Senior	Soph	Senior	Soph	Senior
Public	3.60 (2.00)	4.48 (2.10)	3.69 (1.88)	4.48 (1.97)	9.40 (4.04)	10.63 (4.24)
Catholic	4.34 (1.92)	5.00 (1.96)	4.59 (1.84)	5.35 (1.74)	11.05 (3.56)	12.10 (3.82)
Other Private	4.32 (2.05	5.34 (2.04)	4.78 (2.00)	5.56 (1.94)	11.28 (4.17)	12.74 (4.14)

Source: [2: Table 3].

which are subtests of the longer tests that were administered, are clearly very short and are not generally used standardized achievement tests. What justifies their use as the measures of cognitive outcomes in the high schools? The Report is virtually silent on this question, except to say that

> these tests do not cover subject matter that is an explicit part of the curriculum in the later years of high school. The mathematics items are all rather elementary, involving basic arithmetic operations, fractions, and only a few hints of algebra and geometry [1:159].

This passage should give pause to any reader who seeks answers to questions (2) and (3). For if the mathematics description is any guide, the coverage of the tests is hardly an explicit part of the curriculum even in the earlier years of high school. They appear to measure the outcomes of elementary education. Thus, a face-value interpretation of Table 1 is that students who attend private schools have had better elementary school achievement than those who attend public high schools. Perhaps the private elementary sector is responsible for this difference, perhaps not; that question is neither asked nor answered in the Report. Perhaps the private high school sector adds more to the elementary school achievement than does the public high school sector, perhaps not; that question is asked in the Report, but may not be answerable with these test data. The R, V, M tests are measures of the inputs into the high schools, not of value added; this has implications for the central issue of selection bias, an issue which we discuss at length later.

There are three sectors, represented in the CHK sample by 894 public

schools, 84 Catholic schools, and 27 other-private schools. The very small number of other-private schools is disturbing because that sector is treated on a par with the other two throughout the Report. The extreme heterogeneity of the other-private schools is remarked on by the authors themselves:

> They include the prestigious schools that are often thought of as the private schools in America, schools that roughly coincide with membership in the National Association of Independent Schools. But they also include a wide range of church-related schools, ... some of which operate on a shoe-string; and they include as well schools that have sprung up in response to school desegregation policies and other unpopular policies in the public schools [1:155].

Furthermore, the randomness of this subsample is compromised by nonparticipation. Although the Report is mute on this, we learn from the design manual [4:9] that of the 38 other-private schools originally drawn, only 23 agreed to participate, a number supplemented by an additional drawing of 4 cooperating schools. We doubt that this sample of 27 schools, nonrandomly drawn from an extremely heterogeneous sector, merits serious consideration. Population heterogeneity and small sample size would be reflected, of course, in large standard errors for the sector statistics.

Now consider the individual students whose test scores enter Table 1. The sample design called for obtaining 36 sophomores and 36 seniors in each school, but the actual sample sizes for the cells in Table 1 will be less than those target numbers suggest. Small schools had less than the target number of students, some students were absent, and others declined to participate. By piecing together information in [4] and CHK's computer output,

we are able to make a rough estimate of the numbers of students who actually took the tests. These are displayed in our Table 2, along with the number of students in the sample [1:A10]. The substantial loss of observations (especially in the other-private sector) is further cause for concern. An obvious question is whether the 21 percent of other-private sophomores who did not take the tests would score as high as the 79 percent who did.

Evidently, some of the most elementary descriptive information about the tests, the primary dependent variables of the entire cognitive-outcome study, is missing. CHK, however, express no reservation in their answer to question (1): Private-sector students have higher cognitive outcomes than do public-sector students. In Table 1, the differences in test scores across the sectors are small in absolute terms, but large relative to levels and standard deviations. CHK on occasion use the increase in average test score from the sophomore to the senior year as a standard for describing sector differences. For example, on the 8-item vocabulary test, other-private sophomores answer one more item correctly than do public sophomores (4.78-3.69 = 1.09), an increment which exceeds that provided by the public schools in moving from sophomore to senior year (4.48-3.60 = 0.88). Remarkably, CHK express no reservation about the validity of "nine-tenths-of-a-question" as an index of <u>two years</u> of educational attainment.

C. Statistical Inference

The critical parameters estimated by CHK are regression coefficients for the relationships represented by our Figures 1C, 1D, and 1F. There are surprising gaps in their presentation of the estimates, most notably with respect to conventional measures of reliability. Neither standard errors

Table 2

Number in Sample and Approximate Number and Proportion Taking Tests

	(1) Number in Sample		(2) Number Taking Tests		(3) Proportion Taking Test	
	Soph	Senior	Soph	Senior	Soph	Senior
Public	26,448	24,891	23,700	21,500	.90	.86
Catholic	2,831	2,697	2,700	2,400	.95	.89
Other Private	631	551	500	450	.79	.82

Sources: Col. (1) from [1:A10]; cols. (2) and (3) calculated by present authors from [4] and computer output provided by CHK.

nor confidence intervals, nor tests of statistical significance are to be found in the CHK Report.

The sample sizes shown in Table 2 are so large, however, that one might be tempted to overlook the gaps on the grounds that everything must be significant. Doing so would be a serious mistake. First, in multiple regression analysis, collinearity among the explanatory variables can produce large standard errors regardless of sample size. Second, the High School and Beyond (HS&B) sample was not a simple random sample. By design, the students were not independently drawn across the high school population, but rather in clusters by school. The effective sample size is consequently less than the number of students, so that nominal standard errors computed on a simple-random-sampling assumption must be adjusted upward. At the extreme, if all students in each school were identical, the effective total sample size would be the number of schools (1000 $\pm$) rather than the number of students (58,000 $\pm$). The Report [1:A8] indicates that nominal standard errors for univariate statistics (e.g., means), should be multiplied by factors of 1.5 to 2.5. The appropriate adjustment factors for regression coefficients have not been determined.

Third, the treatment of missing values merits attention. The HS&B sample contains many non-responses on individual questionnaire items. For example, our reading of [4:8-97, 8-204] is that 15% of the students did not report family income, and 20% did not report father's education. In a regression equation with many explanatory variables, the number of complete observations -- those who have no missing values at all -- must be well below the nominal sample size. Jay Noell [5], who ran test-score regressions with HS&B data using a long list of background variables simi-

lar to CHK's, reports that the number of complete observations in the full sample is about 23,000, which is 40 percent of the original 58,000. Douglas Willms [6] used a shorter list of background variables and was able to retain about 50 percent of the nearly 30,000 sophomores.

These considerations suggest that for CHK the task of obtaining appropriate measures of reliability was not trivial. The Report contained no information as to how missing values were handled; their article [2:10] indicates that the pairwise deletion method was used. The Report contained no standard errors at all; their article [2] now presents standard errors for various derived statistics. Those standard errors are computed on a random-sampling assumption: they take no account of clustering, and no account of the implicit reduction in sample size associated with missing values. Consequently they understate true standard errors. As a rough guide which may serve until proper measures of reliability are calculated and reported, we suggest that readers use a $\pm$ 3-sigma (or $\pm$ 4-sigma) rule, rather than the usual $\pm$ 2-sigma rule, in assessing statistical significance. In any event, we have found no indication in the Report that the authors' interpretations and conclusions were arrived at by applying conventional criteria of statistical inference.

As a final note on sample size, consider the number of blacks in the other-private sector of the HS&B sample, a number unobtainable in the Report. Our reading of the computer output is that the entire HS&B sample of 58,000 students contains just 41 blacks in the other-private sector. It is remarkable that CHK had no hesitation in calculating for the Report the number of blacks who would shift to the other-private sector in response to a universal $1,000 increase in income [1:38], a calculation from which they infer the response to tuition tax credits and school vouchers [1:68-73,

230-231]. The inference is, of course, untenable; see Catterall and Levin [7]. Nor has the small sample size for blacks inhibited CHK's publication of a racial segregation index for the other-private sector [1:44; 2:Table 2], an index which, one now sees measures the distribution of just 41 persons across just 27 schools.

3. SECTOR EFFECTS ON TEST SCORES

A. Background Control

There are various ways of presenting the differences in mean test scores across sectors and across grade levels. CHK in effect proceed as follows. Let $\bar{y}_{ij}$ = mean test score in sector i at grade j, with i = 1,2,3 indexing the public, Catholic, and other-private sectors, and j = 1,2 indexing the sophomore and senior grade levels. Then define

$\bar{y}_{21} - \bar{y}_{11}$ = Increment (at sophomore level) for Catholic sector

$\bar{y}_{31} - \bar{y}_{11}$ = Increment (at sophomore level) for other-private sector

$\bar{y}_{12} - \bar{y}_{11}$ = Senior increment in public sector

$(\bar{y}_{22} - \bar{y}_{21}) - (\bar{y}_{12} - \bar{y}_{11})$ = Extra senior increment for Catholic sector

$(\bar{y}_{32} - \bar{y}_{31}) - (\bar{y}_{12} - \bar{y}_{11})$ = Extra senior increment for other-private sector.

For each of three tests, these observed increments, or unadjusted mean differences, as taken from CHK [1: Table 6.1.3, 154; Table 6.2.1, 171; 2: Tables 3, 4], are set out in the first column of our Table 3.

The issue now is the extent to which these observed increments survive statistical control for initial differences among the students entering the several sectors and grades. As CHK put it:

> [T]he differences may well be due merely to the differential selection of different students into the different sectors...

Table 3

Various Measures of Sector and Grade Effects on Test Scores

	Observed	Controlling For				
		Background			Track	Dropout
	(1)	(2)	(3)	(4)	(5)	(6)
READING						
Sophomore Increment						
1) Catholic	.74	.32	.23	.26	.18	-
2) Other-private	.72	.14	.06	.02	-	-
Senior Increment						
3) Public	.88	.73	.75	-	-	.47
4) Extra, Catholic	-.22	-.07	-.13	-	-	0
5) Extra, Other-private	.14	.27	.22	-	-	.34
VOCABULARY						
Sophomore Increment						
1) Catholic	.90	.36	.40	.41	-	-
2) Other-private	1.09	.33	.37	.31	-	-
Senior Increment						
3) Public	.79	.63	.70	-	-	.41
4) Extra, Catholic	-.03	.19	.04	-	-	.20
5) Extra, Other-private	-.01	.17	.04	-	-	.21
MATHEMATICS						
Sophomore Increment						
1) Catholic	1.65	.58	.35	.46	.32	-
2) Other-private	1.88	.56	.32	.22	-	-
Senior Increment						
3) Public	1.23	.88	1.02	-	-	.38
4) Extra, Catholic	-.18	.01	-.02	-	-	.30
5) Extra, Other-private	.23	.17	.14	-	-	.60

Sources:

Col. (1): Rows 1, 2, 3 in [2:Table 4]; Rows 4, 5 are authors' calculations from [2:Tables 3,4].
Col. (2): Rows 1, 2, 3 in [2:Table 4]; Rows 4, 5 are authors' calculations from [2:Table 6].
Col. (3): Authors' calculations from [1:A12-A14].
Col. (4): From [2:15].
Col. (5): Authors' calculations (available on request) from [6].
Col. (6): Authors' calculations from [2: Table 5].

> Are the differences entirely due to selection, or are there also different effects...? [W]hat would be the differences in outcome if the students coming into the different sectors were alike? [1:167].

As their first statistical control, CHK take a set of seventeen background variables drawn from the students' responses on the hour-long questionnaire filled out in class [1:172; 2:9].

At this stage, a conventional analysis-of-covariance approach would begin with the regression, across all students, of test score upon the background variables and a set of dummy variables capturing the sector-by-grade classification. That is:

$$(1) \qquad \hat{y} = x'b + \Sigma_{i=1}^{3} \Sigma_{j=1}^{2} a_{ij} z_{ij},$$

where $\hat{y}$ = fitted test score, x = background vector, and z_{ij} = 1 or 0 according to whether the student is or is not in sector i at grade j. (Throughout our article, "vector" denotes a column, and the prime denotes its transpose; thus $x'b$ should be read as the sum of products of the elements in x with the corresponding elements of b.) The slope vector b is taken to be the same for all sectors, while the intercepts a_{ij} represent the "main effects" of sector and grade. The differences ("contrasts") among the a_{ij} would serve as tentative estimates of adjusted increments to be compared with the corresponding observed increments. If the focus were on sector effects at each grade level, the next step in a conventional approach would separate out the grades and fit a pair of regressions, still running across all sectors:

(2a) $\hat{y} = x'b_1 + \Sigma_{i=1}^{3} a_{i1} z_{i1}$ (sophomores)

(2b) $\hat{y} = x'b_2 + \Sigma_{i=1}^{3} a_{i2} z_{i2}$ (seniors) .

Here the slope vector is allowed to differ by grade level, while contrasts among the a_{ij} again estimate adjusted increments at each grade level.

Neither formulation (1) nor (2) is introduced in the Report, which works with the sample split into two sectors -- public and private (combining sectors 2 and 3) -- and two grades. Four separate regressions (2 sectors x 2 grades) are fitted, with dummy variables capturing the Catholic/other-private dichotomy within the private sector:

(3a) $\hat{y} = x'b_{11} + a_{11}$ (public sophomores)

(3b) $\hat{y} = x'b_{21} + a_{21}z_{21} + a_{31}z_{31}$ (private sophomores)

(3c) $\hat{y} = x'b_{12} + a_{12}$ (public seniors)

(3d) $\hat{y} = x'b_{22} + a_{22}z_{22} + a_{32}z_{32}$ (private seniors) .

In this formulation, the impact of background on test scores is allowed to differ between the public and private sectors ($b_{1j} \neq b_{2j}$), but not between the Catholic and other-private subsectors ($b_{2j} = b_{3j}$).

These twelve regressions (2 sectors x 2 grades x 3 tests) are the only ones tabulated in the CHK Report. The intercepts, slopes, and multiple R^2's are given along with the means of the background variables in [1:A12-A14]. No standard errors are reported for the coefficients, so we cannot assess the plausibility of the point estimates of the b's and a's. For example, is high family income "really" conducive to low test scores, as 5 of the 12 equations say? Do other-private schools really depress test scores relative to Catholic schools, as 4 of 6 equations say? Without standard errors, readers

are unable to judge whether the reported sector differences in background effects ($b_{2j} - b_{1j}$) are real, or merely attributable to chance variation. Indeed, with no information on the fit of the simpler specifications (1) and (2) above, there are no grounds for judging whether the non-additive specification (3) provides a meaningfully better fit. This failure is compounded by the authors' persistence in taking all point estimates literally -- that is, treating them as perfectly reliable -- except when sampling variability provides a convenient rationalization for anomalous results [1:41, 45, 46, 160, 177, 201, 203].

Having fit equations (3), CHK face a mechanical problem in developing estimates of adjusted increments. The sophomore increments for the private sector, for example, measure the vertical distance between the lines (3a) and (3b). Since the lines are not parallel, the answer depends upon the value of x at which the distance is to be measured. CHK take the average public school sophomore as the reference point. Formally, let $\bar{x}_{ij}$ = mean value of vector x in sector i at grade j, and p_{ij} = predicted mean test score for sector i at grade j. From their equations (3), CHK calculate

$$(4) \qquad p_{ij} = \bar{x}_{11}'b_{ij} + a_{ij},$$

(with $b_{3j} = b_{2j}$) and process these predicted means into adjusted increments in the same manner that the observed means were processed into observed increments (described at the beginning of this section). For example, the estimated adjusted increment at the sophomore level for the Catholic sector is $p_{21} - p_{11}$. The results -- which are the content of [1:171, Table 6.2.1 & 175, Table 6.2.2; 2: Tables 4 & 6] -- are displayed in the second column of our Table 3.

CHK read these numbers as showing substantial differences to be attributed to the private sector [1:173-174; 2:11]. In our experience it is customary to cross-validate distance estimates from non-parallel regressions by using alternative reference points. CHK have not done so, but the tabulations in [1:A12-14] do permit us to use the average private sophomore (a mixture of $\bar{x}_{21}$ and $\bar{x}_{31}$) as the reference point to obtain p_{ij}'s. Doing so, we obtain the third column of Table 3, which gives a picture generally less favorable to the private sector.

Still less generally favorable are the estimates from fitting an additive specification introduced for the first time in the CHK article, which shows the a_{ij} estimates from equations (2). These directly give estimated increments at the sophomore level, which we enter in the fourth column of Table 3. The remainder of the column cannot be completed from their article because the model does allow interaction by grade level; but our rough calculation (details available on request) shows that private sector increments at the senior level obtained from (2) are uniformly and substantially less than those from CHK's (3). CHK [2:15] dismiss estimates of additive sector effects from (2) as "inferior," on the grounds that one must allow for a full set of interactions between z and x. The conventional evidence for such a claim would be a demonstration that equations (3) provide a significantly better fit than equations (2), but as we have previously noted, CHK offer no such evidence.

The alternative estimates of background-controlled sector effects spread across the rows of columns 2, 3, and 4 of Table 3 span a moderately wide range. As we shall see, all of these estimates are upwardly biased.

B. Selectivity Bias

CHK use 17 measured background variables to control for initial differences among the students entering the several sectors. How adequate is this control? We doubt that all 17 variables together can substitute for direct initial measures of cognitive achievement, such as would be provided by accurate R, V, and M test scores obtained just prior to entering high school. Consequently, a major selectivity bias problem appears, which may be conceptualized as follows. Suppose that among students of the same measured background, it is the initially higher-scoring students who choose the private sector. Then the omission of those initial test scores, as in CHK's analysis of covariance, would produce a selectivity bias in favor of the private sector. The problem is compounded here because CHK's outcome measures are themselves measuring pre-high-school achievement rather than the outcomes of high school experience.

CHK are well aware that in nonexperimental situations, all versions of analysis of covariance are subject to skepticism on the grounds that the covariates may not capture all relevant preexisting differences. In their words,

> [T]here may very well be other unmeasured factors in the self-selection into the private sector that are associated with higher achievement [1:224].

But they are of several, contradictory, minds about the efficacy of their measured background variables as controls for initial differences. Thus CHK cite the known difference in motivation for education between parents who send their children to private schools and those who send their children to public schools, and flatly assert:

> [T]his difference between parents, by its very nature, is not something on which students in public and private schools can be equated. Thus this [background-regression] approach is a a particularly defective one in comparing public and private schools [1:168].

By the end of the Report, their verdict is altered:

> A large number of background characteristics is introduced, to insure that the selectivity-related differences are controlled for [1:220].

How much insurance, one wonders, can a particularly defective method buy? Soon after the Report was issued, Coleman told an interviewer that by the use of background controls,

> bias resulting from self-selection was minimized... "If anything, we probably overcompensated for the self-motivation factor" (New York Times, April 26, 1981: Y-19).

With underadjustment becoming correct adjustment becoming overadjustment in three steps, it is evident that one cannot rely on CHK's assessments of their own method.

Controlling for prior cognitive achievement would be the most natural approach to obtaining unbiased estimates of high school effects. An alternative approach would focus on the selection process itself. By modeling, and eventually statistically controlling for, the systematic determinants of sector choice, one can estimate the net effect of sector upon outcome by reliance on the remaining sources of variation in sector choice: see Barnow, Cain, and Goldberger [8]. The two approaches are related, and indeed in our

subsequent discussion we will not always distinguish between background variables as influences on prior achievement and as influences on sector choice. Omitting a background variable which is correlated with initial achievement, but not with sector choice, should produce no bias in estimating sector effects. Similarly, omitting a background variable which is correlated with sector choice, but not with prior achievement, should produce no bias in estimating sector effects. Thus full statistical control over either initial achievement or sector choice would suffice.

Consider now the specific set of seventeen variables which constitute CHK's background vector. The list omits prior cognitive achievement, contains poorly measured background variables, and is far from comprehensive. Family income, for example, is obtained by asking the students to guess, in the middle of an hour-long questionnaire, the dollar bracket into which their family's income fell. The student's sex and parents' occupations are on the questionnaire but are not included in CHK's list.

The list includes several variables (e.g., possession of an encyclopedia and of a pocket calculator) which are "not clearly prior" to high school achievement. CHK claim [1:170-171] that inclusion of such variables "overcompensates" for pre-existing differences, and thus tilts the balance in favor of the public schools. But this claim is unfounded. High test scores might lead to purchase of an encyclopedia, and the private sector might produce higher test scores. But it is far-fetched to presume that this chain of causation overrides the direct role of encyclopedia ownership as a proxy for unmeasured family background, and offsets other neglected prior differences among students entering the several sectors.

Among the included background variables, only the two items referring to each parent's desires about the student's college plans directly measure

parental educational motivation. As for the student's own motivation, nothing in the background list bears directly on academic vs. vocational preferences. Private schools appear to offer little in the way of vocational training, as is clear from data in the earlier chapters of the Report [1:79-80, 93, 97]. So vocational preferences, which presumably affect performance on cognitive achievement tests, may well influence sector choice as well. We expect that among students of comparable background, those who are oriented toward vocational and general curricula are more likely to be enrolled in the public sector.

Sector choice and test scores are also likely to be influenced by such diverse and unrelated special disadvantages of students as mental or physical handicaps, or foreign-language-speaking homes, none of which is included in the list. Indeed, none of the above-mentioned factors, apart from parental motivation, is discussed, let alone analyzed, in the Report in connection with the decision to enter a particular sector. There is no discussion of the school administrators' admission and retention policies. We may presume that public schools are the least restrictive, but CHK do not analyze or discuss the private schools' criteria.

The Report is totally silent on the several curricula -- academic, general, and vocational -- taken by the students, although this information was on the questionnaire, and track is a variable that relates to both cognitive achievement and sector choice. From Lutz Erbring's report [9], prepared at the National Opinion Research Center, and dated September 1980, it was evident that in the HS&B sample, the average academic-track public school student scored at about the same level as the average private school student. From computer output now available to us, we learn that the

distribution of students by track differs drastically across the sectors. According to the students' self-reports, the academic/general/vocational mix is 20/46/22 in the public sector, 61/32/6 in the Catholic sector, and 57/37/5 in the other-private sector, at the sophomore level. At the senior level the corresponding figures are 34/38/26, 69/21/9, 70/21/7. Thus it is reasonable to conjecture that had the student's track status been included as a covariate -- to capture initial abilities, proclivities, and interests -- then the private-sector adjusted increments would have fallen substantially.

As we say, CHK's publications [1; 2; 3] are silent on the track variable. Queried just after issuance of the Report, Coleman

> rejected those who faulted him for comparing students in non-academic programs in public schools with those in private schools, where a higher proportion are in academic programs. "The program you are in is not a 'background' characteristic for which you should control statistically," he said. "It has a lot to do with school policies" [New York Times, April 26, 1981: Y-19].

We would agree with the principle that track status is an inappropriate control variable when it is determined by sector policy, imposed as it were on otherwise identical students, in which case it should be viewed as one of the methods by which the sectors produce cognitive achievement. By the same token, it is an appropriate control variable when it is predetermined in the sense of reflecting initial student characteristics. A reasonable position is that it is a mixture of both. Coleman takes a polar stance, one which tilts the CHK study toward overstating the private sector effect on test scores.

In any assessments of Coleman's stance and the Report's neglect of track status, the following considerations are relevant. The variable in question is the student's own track status, as distinguished, say, from an index of whether or not the school offered each of the tracks. The latter variable might be construed as a pure policy, but not the former. Further, Coleman's ex post rationalization in effect views the general and vocational curricula as having no other function than cognitive development. Readers who believe that other functions are also served, such as preparation for life careers, would have been helped by estimates of cognitive outcome differences by track, as an index of the sacrifices made in pursuit of the other objectives.

In this connection, it is instructive to learn from Peng, Fetters, and Kolstad [10:ix] that the HS&B

> study's primary purpose is to observe the educational and occupational plans and activities of young people as they pass through the American educational system and take on their adult roles.

In any event, readers of the CHK Report are ill-served by the omission of all information on the empirical association between track and test score, information which, if present, they could have interpreted in the light of their own judgments of exogeneity and endogeneity.

Some indications of that empirical association in the HS&B sample are now available. Willms [6] analyzed sophomore scores on the full reading and mathematics tests (18 and 38 items, respectively). Because of numerous missing values on background variables and the small sample sizes in the private sector, Willms decided to use a short list of only five background

variables to compare the three tracks in the public and the Catholic sectors. Furthermore, because the students' self-reports of their track status gave proportions which differed substantially from the aggregates reported by the HS&B school administrators, he decided to reclassify the students. For this purpose, students who reported themselves as on the general track, but who also reported that they planned to attend a four-year college or university immediately following high school, were reassigned to the academic track. Thus Willms's academic-track category is properly interpreted as a college-bound category. This reassignment affected about one-quarter of the general-track students in the public sector, and about one-half of the general-track students in the Catholic sector.

Willms fitted separate regressions by track, with dummy variables introduced for sector. His sample sizes were approximately: academic track -- 6000 public, 1200 Catholic; general track -- 4200 public, 200 Catholic; vocational track -- 2400 public, 100 Catholic. To summarize the results, we have converted his estimates, which are in long-test units, into units for the short tests, so that they are comparable to those we have been discussing (details available on request.) On that understanding, Willms's estimates of the Catholic sector increment at the sophomore level by track are: reading -- .12, .36, and .18 for the academic, general, and vocational tracks, respectively; mathematics -- .00, 1.13, and .50 for the academic, general, and vocational tracks. Only the general-track increments are reported to be significant, and as Willms's notes, his standard errors are understatements because they neglect the clustered design.

It is instructive to combine Willms's track-specific effects into an overall estimate of the Catholic sector effect. For illustrative purposes

only, we do so by constructing a weighted average, using a mixture of public sector and Catholic sector weights. (Details of our calculation are available on request.) We enter these in column 5 of our Table 3. Observe the further shrinkage: the Catholic-sector increments at the sophomore level are, for reading, .18 compared with CHK's .32; and for mathematics, .32 compared with CHK's .58.

Other evidence for the absence of a net sector effect on test scores among students on the same track is available in a staff memorandum prepared at the National Center for Education Statistics [11]. Interestingly enough, the HS&B tests analyzed there are two which tap material that is in fact taught in high schools, namely "Science" (for sophomores) and "Mathematics II" (for sophomores and seniors). In the NCES memorandum mean scores for academic-track students are tabulated for cells defined by two sectors (public and Catholic), three ethnic groups (white, black, Hispanic), and three socioeconomic status (SES) categories. We have constructed a weighted average across ethnicity and SES for each sector. We find that among sophomores, the Catholic sector has a positive increment for mathematics and essentially a zero increment for science; among seniors, the Catholic sector has a negative increment for mathematics. (Details available on request.)

The Report does not use, or mention, the widely used econometric approach to selectivity bias in nonexperimental data, the gist of which is a multi-equation model in which the outcome equation is supplemented by an equation determining selection (i.e., sector choice). This model explicitly captures the possibility that outcome and selection may have common measured and unmeasured determinants. Under restrictive conditions, estimation of the multi-equation model will provide unbiased estimates of the net sector effect. At a minimum, it will provide some guidance as to the extent to

which test-score differences might be attributable to preexisting differences rather than to sector effects. In fact, the first step of the econometric approach, namely a probit regression of a student's sector status on measured background, would by itself provide an informative summary of the measured background differences among students entering the sectors. The approach was developed several years ago by the economist James Heckman [12], a colleague of Coleman, Hoffer, and Kilgore at NORC.

Noell [5] has applied several versions of this approach to the HS&B data, using the full reading and mathematics tests for both grade levels. His first-step probit regressions indicate strong effects of background upon sector choice, with region, religious background, and the students' college expectations as of 8th grade -- variables not included in CHK's background vector -- being among the most significant. His second step results -- the adjusted regressions of test score on background -- are mixed. The private sector effect becomes negative at the senior level, but becomes more positive at the sophomore level. The magnitude of these estimated effects is sensitive to the specification of the test-score equation, specifically to the inclusion or exclusion of Catholic religious background as an explanatory variable. Evidently, this first application of the Heckman approach to the HS&B data set has not provided a definitive resolution of the selectivity-bias issue. At this stage, it appears that "strong" conclusions from the HS&B data set are not robust across plausible changes in model specification.

In empirical applications of the econometric approach, two persistent problems are (i) the specification of "exclusions" for the outcome equation, and (ii) high collinearity when such exclusions are not imposed. (See

Barnow, Cain, and Goldberger [8]). Here the exclusions refer to variables which affect sector choice, but do not affect test score. If such variables are picked out on an ad hoc basis, or capriciously, the estimates of sector effect will not be valid. On the other hand, if no exclusions are imposed, the variable constructed in the probit-regression step, as a function of the measured background variables, is likely to be highly collinear with those background variables, when they all reappear in the test-score equation. If so, estimates of sector effects will be unreliable. Others who follow Noell's lead in using the econometric approach may find that the HS&B data set does not contain a rich enough set of measurements. If so, the selectivity-bias issue will remain unresolved.

C. Background Revisited: The "Common School"

CHK [1:176-180] digress from their main focus on the mean differences across the sectors to investigate the interaction between sector and the students' backgrounds. This digression is worth some attention for it led the authors to a statement that was emphasized in press coverage:

> Altogether, the evidence is strong that the Catholic schools function much closer to the American idea of the "common school," educating children from different backgrounds alike, than do the public schools ... Catholic schools more nearly approximate the "common school" ideal of American education than do public schools, in that the achievement levels of students from different parental educational backgrounds, of black and white students, and of Hispanic and non-Hispanic white students are more nearly alike in Catholic schools than in public schools [1:221, 232].

What is the statistical basis for this statement? Regression of test score upon measured background by sector is again the mode of analysis, despite the authors' previous dictum that this method is "particularly defective." But now background is confined to only 5 of the 17 variables, namely family income, father's education, mother's education, race (black vs. white), and ethnicity (Hispanic vs. non-Hispanic), while all three sectors are distinguished.

For each test, 6 regression equations (3 sectors x 2 grades) are fitted:

$$(5) \qquad \hat{y} = x'b_{ij} + a_{ij} \qquad (i = 1,2,3;\ j = 1,2)$$

where $\hat{y}$ = fitted test score, x = background vector (5 elements), the b_{ij} are slopes, and the a_{ij} are intercepts. CHK [1: Table 6.2.3, 178] present selected elements of the b_{ij} estimates, namely the coefficients on the race dummy, the ethnicity dummy, and on a combination of the parental education variables. (The coefficients on income are not given, nor are the race and ethnicity coefficients for the other-private sectors, the latter "because the numbers of blacks and Hispanics in the sample of these schools is small enough to make estimates unstable.") At the sophomore level, the Catholic-sector coefficients b_{21} are smaller (closer to zero) than the corresponding public-sector coefficients b_{11}. Thus, for example,

> The achievement of blacks is closer to that of whites...
> in Catholic schools than in public schools [1:178].

Also, in the Catholic sector the senior-level coefficients are generally smaller than the corresponding sophomore-level coefficients, while the

reverse is true in the public sector. Thus,

> [N]ot only is achievement more alike among students from different backgrounds in the Catholic schools than in the other sectors, it becomes increasingly alike from the sophomore to the senior year [1:179-180].

This then is the empirical source of the "common school" conclusion.

It is hard to treat this material seriously. First, discarding 12 of 17 explanatory variables may be expected to distort the estimates of the remaining 5 coefficients. The potential for bias would be indicated by a comparison of R^2's from regressions using the long and short lists of background variables. The Report is silent, but CHK's computer output indicates that discarding the 12 variables reduced the multiple R^2's by about one-third and substantially changed regression coefficients on the retained variables. Second, reported differences between public-sector and Catholic-sector coefficients might be attributable to chance variation.

A third line of argument suggests that the smaller slopes in the Catholic sector are an artifact of selection. A general result in regression analysis is that selection on the dependent variable tends to attenuate slopes. And our previous discussion has already indicated that the Catholic sector appears to be selective, in the statistical sense, of students with favorable measured background and unmeasured academic abilities. (A similar situation will arise if one considers the public school academic track as analogous to a private school. One would expect to find flatter relationships of test score on background within academic track than across all tracks.) More pointedly, the selectivity argument suggests that students with unfavorable measured backgrounds who nevertheless enter the private

sector are precisely those whose unmeasured academic proclivities were unusually favorable, which is a restatement of the flat-relationship result. (For a formal discussion of conditions under which selection guarantees attenuation of slopes, see Goldberger [13].)

On balance, it is evident that CHK's "common school" conclusion has no solid empirical support.

D. Senior Increments

To assess sector effects on senior test scores, CHK [1:167-169] introduce sophomore test scores as an alternative to background measures for the statistical control. More precisely, they suggest that sophomore-to-senior change, $y_2 - y_1$, serves as a measure of outcome produced by the schools. A plausible model at this point might be

$$(6) \qquad y_2 = \beta_o y_1 + x_1'\beta_1 + x_2'\beta_2 + \Sigma_{i=1}^{3} \alpha_{12} z_{12} + u,$$

where y_1 and y_2 are the sophomore and senior test scores, x_1 and x_2 are the sophomore and senior background vectors, u is a residual, and the α_{12} are intercepts (whose differences would be read as sector increments). In the light of (6), CHK's direct use of change scores as estimates of the α_{12} requires the stringent assumptions that $\beta_o = 1$ and $\beta_1 = \beta_2 = 0$. CHK [1:169] remark on the first of these, and suggest [2:12] that ruling out an interaction between level and change (that is, imposing $\beta_o = 1$) tends to bias their results in favor of the private sector. (The rationale apparently is that $\beta_o > 1$, and y_1 is higher in the private sector. Why β_o should be greater than unity is not clear to us.) On the other hand, they do neglect background entirely in this phase, suggesting [2:12] that the control is not necessary. It seems to us that omitting x_1 and x_2 does bias the

results in favor of the private sector. For example, a family breakup producing a single-parent household might occur more frequently among seniors in the public sector and concurrently reduce achievement. If so, by ignoring the initial and subsequent levels of x, CHK's approach would tend to penalize the public sector for the decline in y_2.

As it happens, equation (6) cannot be fitted at the individual student level in the HS&B sample. Neither y_1 nor x_1 is available for individual seniors, because the sophomores and seniors are not the same persons. (This limitation will be removed when the second wave of the HS&B survey returns to the sample schools two years after the first wave.) Equation (6) could have been fitted at the level of school averages, but CHK proceed directly to the sophomore-to-senior gains, averaged by sector [1:180-185; 2:11].

The observed average gains are in fact the observed mean differences $\bar{y}_{i2} - \bar{y}_{i1}$ (i = 1,2,3) which were previously processed into the increments given in our Table 3 (lines 3, 4, 5 of the first column). Notice that the "extra senior increment" for the Catholic sector, relative to the public sector, is negative for all three tests. At this point, CHK call attention to a particular selectivity bias problem with the $\bar{y}_{12}$. The senior test score distributions are distorted by the presence, or rather the absence, of dropouts. From the rosters of the sampled schools, they estimate that as a consequence of students having dropped out of school during the last two years of high school, 31% of the senior class is missing in the public sector, 13% in the Catholic sector, and 15% in the other-private sector. (These rates are too high. Official NCES statistics [14:15] indicate that the national dropout rate is 20%, which is well below the national

figure implied by CHK's sector rates, namely 29%.) Presumably it is the lower-achieving student who tends to drop out, so the observed $\bar{y}_{12}$ overstate the means for the full senior class, the overstatement being largest in the public sector where the dropout rate is highest.

CHK's solution to this serious problem is to produce for each sector a mean for the hypothetical full senior class, y^*_{12}, and calculate dropout-adjusted sophomore-to-senior changes, $y^*_{12} - \bar{y}_{11}$. These are given in the upper panels of CHK's Table 6.2.5 [1:184] and Table 5 [2]. Processing them into increments gives the sixth column of our Table 3.

The public sector effect ("senior increment") is reduced substantially as compared with the observed figures in column 1, and even as compared with the background-adjusted figures in column 2. The Catholic sector effects ("extra senior increments") are now non-negative, and the extra senior increments for the other-private sector have become more positive. It is just these figures which led CHK to

> the conclusion that greater cognitive growth occurs between the sophomore and senior years in both private sectors than in the public sector [1:225],

reversing their earlier view that the observed growth seemed "very small everywhere" and "very much the same among the different sectors" [1:158].

Upon examination, it was faulty methodology that generated their findings. CHK describe their calculation of the hypothetical y^*_{12} as being based on the assumption that

> the dropouts came from the lower 50 percent of the test score distribution on each test and were distributed in that lower half in the same way that remaining seniors in the lower half

are distributed. What this means in effect is that within the lower half of the senior test score distribution, and within the upper half, the distributions do not change; but the lower half, augmented by the dropouts, becomes a larger share of the total.

This assumption probably errs on the side of being favorable to those schools with high proportions of dropouts (in this case public schools), because dropouts are probably concentrated more toward the bottom of the distribution than is assumed. Thus the assumption is probably conservative with respect to the inference at hand: that is, the greater achievement growth in the private sector [1:182-183].

We confess to some uncertainty as to the arithmetic operation being described. Our best guess is that y^*_{i2} is the score which 100 t_i% of the observed seniors exceed in sector i, where

$$(7) \qquad t_i = \frac{1}{2(1-w_i)},$$

with w_i being the dropout rate for sector i. A rationale for this estimator would run as follows. Suppose that the full population test score distribution were symmetric, so that half were above the mean. If the proportion w_i of that population drop out, all from the lower half, then the upper half of the original group would constitute the proportion t_i of the surviving seniors. With $w_1 = .31$, $w_2 = .13$, $w_3 = .15$, the upper half of the original groups would constitute $t_1 = .72$, $t_2 = .57$, and $t_3 = .59$ of the surviving seniors in the public, Catholic, and other-private sectors respectively. Thus y^*_{i2} is taken to be the score such that 72% of the observed

public seniors score above it, etc.

If this is CHK's estimator, then they did not at all err on the side of being favorable to the public sector. For the validity of the construction above as an estimator of the mean rests only on the assumption that dropouts came entirely from the lower half of a symmetric distribution, and not on whether they were uniformly distributed across that lower half or more concentrated toward the lower tail.

More important, it is evident that CHK are adjusting the wrong distributions in their dropout procedure. To the extent that dropping out is determined by background, it is the conditional distribution of senior test score given background, rather than the unconditional distribution, that demands adjustment. Recall that the background controls have been dispensed with in this portion of the CHK analysis. Surely, the measured background factors -- family income, parental education, both-parents-present, ethnicity, and the parents' thoughts on whether the student should go to college -- are predictive of dropping out, as well as of senior achievement. CHK have, in their change measure, reintroduced the very background differences they had been so insistent on controlling for in the regression phase.

A formal specification and estimation of the dropout process might be based on the econometric approach mentioned earlier. Without it we have no definitive figures for the sector effects, if any, which would remain after controlling both for background and for dropping out. But let us recall that CHK's initial 17-variable background regressions already provided adjusted measures of sophomore-to-senior growth in the several sectors, given in rows 3, 4, 5 of the second column of our Table 3. To the extent that dropping out does depend on background, those figures can now be viewed

as controlling both for background's direct effect on outcomes and for its effect on dropping out. Indeed, CHK could agree with this notion, for they wrote:

> [F]amily backgrounds of seniors are slightly higher than those of sophomores, a difference that is attributable to greater dropout rates ... for students from lower backgrounds [1:173].

According to the second column of Table 3, the public sector has larger growth than the Catholic sector for two of the three tests, while growth in the other-private sector is moderately greater than in the public and Catholic sectors. In the Report, CHK [1:175] caution against using those figures because of their neglect of the dropout phenomenon. The figures they preferred were the dropout-adjusted ones, which we now see are biased. In their article, CHK appear less certain: the background-adjusted extra senior increments are now introduced after, and as a "variant" on, the dropout-adjustment method of assessing differential gains, and are said to suggest that the "dropout correction may have been too great" [2:14-15]. Nevertheless, a paragraph or so later their reservations are abandoned, and they return to the strong language of the Report: "considerably greater" and "substantial" differences in favor of the private sectors.

The background-controlled increments are arguably preferable to the dropout-adjusted increments, but selectivity bias remains as a confounding factor. As CHK remark, since

> dropping out of school is an act of negative selection, the students who drop out are very likely lower achieving than those from similar backgrounds who remain in school [1:181].

The background-controlled increments would seem to be more upwardly biased in the public sector if it were to turn out that this sector has a higher dropout probability for students of comparable measured background. (As matters now stand, there is no information on this issue in the CHK reports.) However, if students were fully comparable in their backgrounds, allowing for the unmeasured as well as the measured determinants, it is not at all obvious that the dropout probability for the public sector would be higher. As a generalization, parents who anticipate that their children will drop out are unlikely to pay the extra expenses required for private schooling, so private-school enrollees have a retention propensity that reflects parental motivations that are like those previously discussed in connection with test-taking abilities. It is conceivable that, after controlling for this retention propensity as well as for the available background variables, the dropout rate in the public sector would be lower, precisely because public school policies may be more lax, thus encouraging retention. In this light, using the background-controlled sophomore-to-senior changes may not be biased in favor of the public sector.

CHK carry out further calculations to convert the senior increments (additional number of test items answered correctly) into annualized growth rates. To the two points y^*_{12}, $\bar{y}_{11}$ for each sector they fit a differential equation, the parameter of which, interpreted as the rate of growth g, is estimated as

$$(8) \qquad \log (1+g) = (1/2) \log ((T-\bar{y}_{11})/(T-y^*_{12}))$$

where T = total number of items on the test (= maximum possible score). These growth rates are the entries in the lower panels of CHK's Table 6.2.5

Table 4

Various Measures of Growth Rates in Test Scores,
Sophomore-to-Senior, by Sector

	Dropout Adjusted			Background Controlled	
	Ceiling Formula		Conventional Formula	Ceiling Formula	Conventional Formula
	Log (1)	Arithmetic (2)	Log (3)	Log (4)	Log (5)
READING					
Public	.06	.05	.06	.09	.10
Catholic	.07	.06	.05	.09	.08
Other-private	.13	.11	.09	.14	.13
VOCABULARY					
Public	.05	.04	.05	.08	.08
Catholic	.10	.09	.06	.12	.10
Other-private	.11	.10	.06	.12	.09
MATHEMATICS					
Public	.02	.02	.02	.06	.05
Catholic	.05	.05	.03	.06	.04
Other-private	.08	.07	.04	.07	.05

Sources:

Cols. (1): Calculated by applying our equation (8) to upper panel of Table 6.2.5 [1:184]. In principle, figures should coincide with those in the lower panel of that table. The slight discrepancies are apparently attributable to rounding.

Cols. (2) and (3): Calculated by applying our equations (9) and (10) respectively to upper panel of CHK's Table 6.2.5 [1: 184].

Cols. (4), (5): Calculated by applying formulas corresponding to our equations (8) and (10) respectively to data in CHK [2:Tables 4,6].

[1:184] and Table 5 [2], reproduced here as the first column in our Table 4. In their formulation, the growth rate is in effect the percentage decrease in the number of incorrect answers. Essentially the same figures emerge from the following arithmetic variant of CHK's formula:

$$(9) \qquad g = (1/2)(y^*_{12} - \bar{y}_{11})/(T-\bar{y}_{11});$$

see the second column of our Table 4. A more conventional growth rate would be defined in effect as the percentage increase in the number of correct answers. Calculating this as

$$(10) \qquad \log (1+g) = (1/2) \log(y^*_{12}/\bar{y}_{11})$$

gives the numbers in the third column of Table 4. The fourth and fifth columns of Table 4 come from applying (8) and (10) respectively to the background-adjusted, rather than dropout-adjusted, increments.

Evidently, the choice of growth rate measure is a nontrivial matter. Columns (3) and (5) in Table 4 are less favorable to the private sectors than their counterparts (1) and (4). CHK's rationale for choosing their distinctive growth rate formula is that it corrects for ceiling effects (high-scoring students are limited in the number of additional correct answers they can get). Actually, their choice is an arbitrary one from among the alternative devices for correcting a deficiency of the test instruments they used, and one which, in the event, tilted the balance toward the private sector.

To summarize our assessment of the CHK analysis of senior increments: We find no evidence of a positive private-sector effect on growth from sophomore to senior year. Contrary to CHK's position, their "extra senior

increments" are contaminated by selection bias, their dropout adjustment is inappropriate, and their growth rate formula is arbitrary. In each instance, their choice has tilted the balance in favor of the private sector.

Our summary stands in sharp contrast to that by CHK:

> The estimated learning rates show great differences between students in other private schools and those in public schools .... It is true that various assumptions are necessary, as discussed earlier, to estimate such rates. But if the assumptions are favorable to any sector it is probably the public sector The evidence is thus rather strong that average achievement growth is considerably greater in the private sectors than it is in the public sector [1:185].

4. SCHOOL POLICIES

A. Claims

We turn now to CHK's final phase, their "third method for studying the differential effects of public and private schools," which they call the "most valuable portion of the analysis" [2:16, 23]. The full sweep of the claims has been conveyed by Coleman:

> If Catholic or other private schools bring about higher achievement for comparable students, and if they do so through those attributes measured in the research [i.e., school policies] which distinguish Catholic and other private schools from public schools, then we should find achievement differences among schools within any sector, public or private.... [W]ithin the

> public sector, the performance of students similar to the average public school sophomore, but with the levels of homework and attendance ... and disciplinary climate and student behavior attributable to school policy in the Catholic or other private schools, the levels of achievement are approximately the same as those found in the Catholic and other private sectors.
>
> The first implication of these results is that they strongly confirm the school-effect results found by the other two methods. For the selection hypothesis necessary to account for these differences must be especially tortured, operating not only between sectors but also to the same degree within sectors, and operating to select students, on the behavior variables indicated above, into schools with particular disciplinary climates... .
>
> A broader implication holds as well: ... these attributes described above are in fact those which make a difference in achievement in all American high schools no matter what sector they are in. Schools which impose strong academic demands, schools which make demands on attendance and on behavior of students while they are in school are, according to these results, schools which bring about higher achievement [3:24-25].

Our discussion will first sketch what CHK did, then explain why their approach was wrong, and, finally, explain why, <u>if</u> correct, their approach would lead to preposterous conclusions.

B. <u>Procedures</u>

The empirical core of this phase of CHK's analysis consists of regressions of test score on background and policy variables, run on the

public sector only, separated by grade level:

$$(11) \qquad \hat{y} = x'b_{1j} + s'c_{1j} + a_{1j}, \qquad (j=1,2)$$

where $\hat{y}$ = fitted test score, x = background vector (17 variables), and s = "school policy" vector (13 variables). Auxiliary to (11) are the regressions of each of the school policy variables upon the background variables, which may be represented in multivariate format as:

$$(12) \qquad \hat{s}' = x' F_{1j} + h'_{1j} \qquad (j=1,2)$$

$$(13) \qquad \hat{s}' = x' F_{2j} + \Sigma_{i=2}^{3} z_{ij} h'_{ij} \qquad (j=1,2).$$

Here the F_{ij} are matrices of slopes, while the h_{ij} are vectors of intercepts. These equations are fitted separately at each grade level, to the public sector (12) and to the (combined) private sector (13). The auxiliary regressions are then evaluated at a common reference point, namely the average public school sophomore's background, to give background-controlled measures of the policy vector as

$$(14) \qquad q'_{ij} = \bar{x}'_{11} F_{ij} + h'_{ij} \qquad (i=1,2,3;\ j=1,2)$$

(with $F_{3j} = F_{2j}$). Converting into increments by

$$(15) \qquad d_{ij} = q_{ij} - q_{1j} \qquad (i=2,3;\ j=1,2),$$

we have the extra level of the policy variables set in each of the private sectors, relative to the public sector, after controlling for measured background. Finally, CHK multiply these policy increments by the policy coefficients in (11) and sum to obtain

$$(16) \qquad r^*_{ij} = \sum_{k=1}^{13} d_{ijk}\, c_{1jk} \qquad (i=2,3;\ j=1,2),$$

where k indexes the 13 school policy variables. In words, the r^*_{ij} represent the predicted change in public sector test scores that would result if the public sector policies were changed to the levels that prevail in private sector i at grade level j -- all after controlling for background.

This is an elaborate procedure. CHK's previous use of regressions (see (3) - (4) above) would require them merely to run equation (11) separately for each sector, evaluate those equations at a common reference point for x, and directly calculate the increments in y for the public sector after assigning private levels of s. CHK introduce (12) - (16) because they recognize that the policy differences are themselves partly attributable to background differences [1:209]. That the policies are not, after all, purely exogenous is an important point.

Indeed their route is even more circuitous than so far described. They first lead the reader through a maze of preliminary analyses [1:198-206] in which test scores are regressed on the CHK "short list" of five background variables and, in turn, several separate "school policy" variables. This maze implies some startling results. For example, one of the CHK school-policy variables, absenteeism, shows an unusually large effect on test scores. Taking the regression results in Table 6.3.2 [1:202] together with the variable coding [1:B3], we calculate that four days of additional attendance per semester would raise mathematics test scores by .85 points. Now, the predicted increase in mathematics test scores between the sophomore and senior years in public schools is .88 (see column (2), Table 3). So one has the implausible conclusion that four days of attendance per semester are worth as much as the last two

years of high school in terms of test score gains. Implausible, that is, if one views the absenteeism variable as a measure of school policy. On the other hand, if absenteeism is viewed as an indicator of student-specific traits -- prior academic failings, dislike of school, etc. -- the finding is entirely plausible.

CHK also report the effect of school size on test scores [1:203-205]. The effect is positive, and public schools are larger than private schools. Since school size is a variable which differs markedly across the sectors, it meets the criterion which CHK say they used in assembling their policy vector, s [1:197]. But school size is also the only policy variable mentioned which is generally favorable to the public sector. And it is not included in their full policy vector, to which we now turn.

There are thirteen variables in their full set of "school policies", all obtained from the student's questionnaire. It is instructive to note that this same set of variables is variously referred to as "school characteristics and student behavioral variables" [1:210], "school functioning" variables [1:213; 2:21], "school factors" [1:206, 207, 212], and "student behavior and school climate" [2:19]. Grouped into five areas, the list is shown in [1: Table 6.3.4; 2: Table 7] and is discussed in some detail in [1:207]. The interested reader should consult the original survey questions from which those items were selected. Doing so will still leave one mystified as to why "disciplinary climate" is measured in part by students' ratings of their "teachers' interest in students."

CHK's approach to jointly analyzing the effects of the thirteen school policy variables upon test scores is represented by our equations (11) - (16). In their article, Table 7 [2:19] displays the d_{ij} of equation (15),

while Table 8 [2:21] displays the r^*_{ij} of equation (16), in terms of components built up separately for each of the five policy "areas". In the Report, those displays appear as Tables 6.3.4 and 6.3.5 [1:210, 213]. The underlying estimated regressions (11) - (13) are nowhere presented. It is evident, nonetheless, that some or all of the coefficients on the disciplinary climate variables in the test score regression (11) were negative, at the sophomore level. That is to say, good disciplinary climate produced lower test scores. Noting this anomaly, CHK [1:216, 2:22] undertake a roundabout subsidiary calculation [1:217-219], which, as far as we can determine, amounts to re-estimating equation (11) after excluding the student behavior variables. The results from this re-estimation are not shown but CHK refer to Table 6.3.6 [1:218] in claiming that most of the re-estimated coefficients on the disciplinary climate variables were positive. They then interpret this result as a demonstration that disciplinary climate operates <u>through</u> student behavior to produce achievement [1:219; 2:23]. It is not clear to us whether this subsidiary causal ordering among their policy areas is or is not valid. What is clear is that it will be incorrect to credit <u>both</u> disciplinary climate <u>and</u> student behavior with positive effects upon achievement.

The results for the final calculations (16) are shown in CHK's Table 8 [2:21] and Table 6.3.5 [1:215]. Their Table 8 is purported to show the achievement "differences within the public sector associated with the behavioral and school differences that remain between private and public schools when student backgrounds are controlled" [2:20]. The quoted passage may be unclear. Their reasoning is that if these increments (shown in the "accounted for" rows of their Table 8) are equal to the previously calcu-

lated background-controlled increments (shown in the "overall" rows of their Table 8), then the test scores of public school students would equal those of their hypothetical counterparts in the private sector, if only they were given the same level of the school policies.

There are 12 pairs of numbers in CHK's Table 8 which show the "total accounted for" and the "overall" differences in test scores. Coleman's claim is that the numbers in each pair are "approximately the same" [3:25]. As the reader can see, at most 6 of the 12 satisfy that description. Consider next the 7 pairs in which the "accounted for" number exceeds the "overall" number. CHK's interpretation should be that the public schools are more efficacious with these school policies than the private schools are. In fact their interpretation is that the analysis is "imperfect" and shows "puzzling differences" [1:214]. Nevertheless, the final judgment of CHK is that "the policy differences affect public achievement just as they do private school achievement" [2:23], which leads Coleman to the sweeping claims we quoted at the beginning of this section.

CHK have not run regressions of test score on policy and background within the private sector. Nor have they offered any evidence that the coefficients on s are significantly different from zero in the public sector.

C. Critique

We advise readers not to devote much energy puzzling over the issues of ambiguity of definition, sector efficacy, or statistical significance regarding the CHK policy variables. The variables do not define, and only remotely reflect, school policies. The variables are the student's description of his or her personal behavior and perceptions of others' behavior.

In contrast to these items are some HS&B questions asked of school officials that were about school policies. But these responses make no appearance in the CHK analysis.

The "school policy" variables used by CHK have all the appearance of being (a) primarily, reflections of student background characteristics not otherwise controlled for; and (b) secondarily, endogenous outcomes reflecting school achievement; and (c) least of all, exogenous school policies. In this light, we see that CHK are attributing to the public schools negative effects that reflect sources (a) and (b).

Consider the variables defined by whether or not the student has taken an advanced mathematics course or an honors course in mathematics or in English. Are these not indicators of academic abilities and motivations? Are these not also measures of scholastic achievement, just as the test scores themselves are? "Having taken an advanced mathematics course" is a variable whose role in a regression equation explaining mathematics test scores is similar to that of a variable defined as "having achieved a high grade in a previous mathematics course." Does the latter represent a school policy? Yes, to some extent. But is it sensible to estimate the effect of school policies with such a variable? (Note that the variable for having taken an advanced mathematics course is not defined by whether or not the school offers the course.)

"Doing homework," "attending school," and "cutting class" are similarly aspects of a student's behavior that, at face value, are predominantly reflections of the student's motivations and academic orientation. Are we to believe that the role the school has played in shaping the student's motivation and orientation justifies calling these variables "school policies" as

soon as their dependence on 17 measured background variables has been extracted? No doubt some of the variation in the CHK school policy variables is explained by the 17 background variables. CHK do not tell us how much, but it does not really matter, because the 13 so-called policy variables continue to reflect other background characteristics of the students. Recall that analysis of covariance on measured background is in their words a "particularly defective" method for controlling for such difficult-to-measure concepts as student motivation and ability.

Indeed, CHK themselves raise the issue that we have been discussing:

> One might argue that ... the kind of students who tend to be lower achievers are those who are absent or cut classes, and it is not the absences themselves that reduce achievement. This may be so, and the issue certainly merits further attention [1:200].

But they dismiss this argument in the very next sentence on the grounds that the regression coefficients of y on s are similar "in the different sectors, where policies lead to very different levels of absenteeism." The force of this logic escapes us entirely.

Two additional wrinkles in the CHK patchwork of school-policy variables need to be discussed. First, CHK construct school means for "cutting class" and for "absenteeism" that omit the student's own responses, and assign these school means to the student [1:215-216]. Evidently, all students in a school are assigned approximately the same value for the class-cutting and the absenteeism variables. CHK would have us believe that this purges the variables of their background component. But a simple example should

dispose of this spurious claim.

Assume y_i = earnings for the ith worker, a_i = the worker's age, and a regression produces $dy_i/da_i > 0$ as evidence that a_i has a positive effect on y_i. Suppose another hypothesis is that a firm's "age policy" has an effect on earnings. To test this, y_i is regressed on $\bar{a}$, where $\bar{a}$ is the mean age of the ith worker's co-workers. It is found that $dy_i/d\bar{a} > 0$. Is this evidence for the hypothesis that a firm's "age policy" has a positive effect on the earnings of a given worker? Surely a counterinterpretation is that $\bar{a}$ is a proxy for a_i. In fact if a_i were a test score, it might be so fallible a measure that $\bar{a}$ could be a better estimate of the true score. A natural way to proceed is to regress y_i on a_i and $\bar{a}$, which will indicate whether $\bar{a}$ has an effect after controlling for a_i. Evidently, CHK would find it sufficient to regress y_i on $\bar{a}$ and some x_i, representing the worker's "background," and then assert that $\bar{a}$ shows the effect of the firms's "age policy."

A second wrinkle introduced by CHK in their quest for a "school policy" interpretation is to rely on the variables that define the student's perception of other students' behavior. They claim that

> disciplinary climate ... and student behavior in the school ... characterize the school as a unit, rather than the student. They are least susceptible to the alternative selection hypothesis, which for them must become especially tortured [2:18].

By this logic, any school-level variable that is constructed by averaging students' characteristics -- e.g., the students' mean score on a mathematics test taken in eighth grade -- is on the same footing as a school variable that

is an actual policy -- e.g., a policy of not giving grades. CHK could say, after all, that both variables "characterize the school as a unit," and, going on the offensive, assert that it would be "especially tortured" to contend that eighth grade test scores reflect initial differences among the students entering the high schools.

With their disciplinary climate variables CHK have shifted attention from school policy variables towards school-level variables of whatever nature. Their use of a student-behavior variable, which includes such behavior as "attacking teachers," presents a different question; namely, Is a student's test score affected by the background characteristics of his or her peers? This is indeed an important issue for many parents, but it does not address the question, What is the effect on a student's test score of school policies? In passing, we might note that an improvement to the CHK specification to measure the effect of the peer group's disciplinary behavior is a regression that also includes as a variable the student's own prior disciplinary conduct. The peer-group effect would generally be biased if the individual effect were not controlled, just as, in the example above, $dy_1/d\bar{a}$ was biased in the absence of a control for a_1.

D. Assessments

In assembling their list of 13 school policy variables, CHK selected some items from the questionnaire, but not others. In particular, in defining the "coursework" variable, they chose (a) the algebra, geometry, and trigonometry courses, but not the calculus, physics, or chemistry courses; and (b) "advanced or honors program" in English or mathematics, but not "remedial programs" in English or mathematics.

We do not know why CHK made these choices. Had they used the items they

passed over, we suspect that they would have reached the following conclusions: (1) students who have taken, or are taking, calculus score higher on the mathematics test, even after controlling for background; (2) students who have taken, or are taking, remedial mathematics courses score lower on the mathematics test, even after controlling for background. Consequently, if CHK were to adhere to their position that "coursework" is a school policy, they would conclude that removing students from remedial math courses and/or putting them into calculus courses would raise mathematics test scores by the amounts shown by the coefficients in their regressions.

Had CHK presented such findings on the calculus and remedial variables, the findings, at least, would have been met with derision. Should the reaction be any different because they are using "advanced math" and "honors" variables, instead?

A final note: Recall Coleman's stance that "academic track" is a policy variable, rather than a proxy for the background abilities and motivations of the students. If so, shouldn't it have been included in the policy vector? Had that been done, another powerful policy would have been revealed: just shift students from vocational and general tracks to academic tracks and secure the gains measured by the coefficients on track! The fallacy involved would be precisely the same as in the CHK interpretation of their school policy variables.

5. CONCLUSION

In our assessment of the CHK analysis of cognitive outcomes we have found no basis for accepting their conclusions and no merit in their analysis. Their principal conclusions about the cognitive gains attributable to the private sector and to certain school policies are not warranted by

their data. Their research methods, or, alternatively, their execution of the methods, are replete with flaws. The presentation of their analyses is confusing, incomplete, and biased. Their mistakes and their style are one-sided -- pro-private. It is as if they decided to write a brief for the proposition that society shift to the subsidization of private schools and away from the subsidization of public schools. These faults are also evident in other aspects of their Report: (a) the relation between private schooling and segregation (see Taeuber and James [15]); and (b) the relation between income and tuition tax-credits, on the one hand, and attendance at, or choice of, private or public schools, on the other (see Catterall and Levin [7] and Goldberger [16]).

CHK conclude on the basis of four phases of analyses that the private sector has beneficial effects on cognitive achievement. We summarize our assessment of these analyses as follows.

1. Their first analysis used multiple regression with 17 background variables. Our discussion, along with that of previous commentators (see Educational Research Service [17] and Murnane [18]), noted the presumptive bias against public shools in this approach. When academic track is used as a reasonable alternative control for this selection bias, the outcomes in favor of the private sector virtually disappear.

2. A second analysis purported to show that Catholic schools are more egalitarian because they produce more similar test scores among students with diverse socioeconomic backgrounds. This approach was twice biased: first, by the exclusion of relevant background variables (12 of the 17); second, we surmise, by the truncation of student cognitive abilities in a narrower range in the Catholic sector than in the public sector.

3. The third analysis was the study of sophomore-to-senior change in test scores. The CHK pro-private conclusion here was due to the absence of control over background. Our measured sophomore-to-senior change, using background controls, showed that the advantage to the private sector disappears (if a conventional growth rate formula is used) or is sharply attenuated (if CHK's ceiling formula is used). The background-control approach is itself not adequate, however, so its results are only indicative of the direction of bias in the CHK approach.

4. CHK's final analysis, in which they attribute the test-score gains in the private sector to selected school policies, strikes us as patently fallacious. School policies get credit for outcomes that are due to student backgrounds. In this section CHK push their pro-private methods and interpretations to fanciful extremes. For CHK, taking an advanced mathematics course and, for Coleman, taking the academic track, are "school policies." And to measure the effects of such "policies" on test scores, they merely read the coefficients of their "policy" variables from regressions that also include the background variables. The pitfalls of selection are apparent. What is true of the variables "advanced mathematics course" and "academic track" is true of "private sector"; namely, they all contain an obvious bias due to omitted student ability and background characteristics.

REFERENCES

[1] James Coleman, Thomas Hoffer, and Sally Kilgore, Public and Private Schools, Report to National Center for Education Statistics, by National Opinion Research Center, Chicago: March 1981. Mimeo.

[2] ___, "Cognitive outcomes in public and private schools." Mimeo. University of Chicago, December 1981.

[3] James Coleman, "Private schools, public schools, and the public interest," The Public Interest, No. 64, Summer 1981, pp. 19-30.

[4] National Opinion Research Center, High School and Beyond, Information for Users, Base Year (1980) Data, Version 1, Chicago: NORC, December 1980.

[5] Jay Noell, "The impact of private schools when self-selection is controlled: A critique of Coleman's 'Public and Private Schools'", Washington, D.C.: U.S. Department of Education, September 1981, Revised. Mimeo.

[6] Douglas Willms, "Achievement outcomes in public and private schools: a closer look at the High School and Beyond data," Stanford University: Institute for Research on Educational Finance and Governance, October 1981. Mimeo.

[7] James S. Catterall and Henry S. Levin, "Public and private schools: evidence on tuition tax credits," Department of Education, University of California at Los Angeles, November, 1981. Mimeo.

[8] Burt S. Barnow, Glen G. Cain, and Arthur S. Goldberger, "Issues in the analysis of selectivity bias," pp. 43-59 in E. W. Stromsdorfer and G. Farkas, editors, Evaluation Studies Review Annual, Volume 5, Beverly Hills: Sage Publications, 1980.

[9] Lutz Erbring, High School and Beyond Summary Report: An Overview of Outcomes in Secondary Education, Report to National Center for Education Statistics, by National Opinion Research Center, September 2, 1980. Mimeo.

[10] Samuel S. Peng, William R. Fetters, and Andrew J. Kolstad, High School and Beyond: A National Longitudinal Study for the 1980's: A Capsule Description of High School Students, National Center for Education Statistics, April 1981.

[11] Bill Fetters, Jeffrey Owings, Sam Peng, and Ricky Takai, "Review of NORC report Public and Private Schools," Washington, D.C.: National Center for Education Statistics: Memorandum, June 28, 1981.

[12] James J. Heckman, "The common structure of statistical models of truncation, sample selection, and limited dependent variables, and a simple estimator for such models," Annals of Economic and Social Measurement, 5, 1976, 475-492.

[13] Arthur S. Goldberger, "Linear regression after selection," Journal of Econometrics, 15, 1981, 357-366.

[14] National Center for Education Statistics, Educational Digest 1980, Washington, D.C.: NCES, 1981.

[15] Karl E. Taeuber and David R. James, "Racial segregation among public and private schools," University of Wisconsin-Madison: Center for Demography and Ecology, November 1981. Mimeo.

[16] Arthur S. Goldberger, "Coleman goes private (in public)," Stanford: Center for Advanced Study in the Behavioral Sciences, May 1981. Mimeo.

[17] Educational Research Service, "Coleman report on public and private schools: the draft summary and eight critiques," School Research Forum, April 1981.

[18] Richard J. Murnane, "Evidence, analysis, and unanswered questions: Coleman's new study, Public and Private Schools," Yale University: Institution for Social and Policy Studies, April 1981. Mimeo.